

## Article

# Using behavior coding to evaluate the effectiveness of dependent interviewing

by Joanne Pascale and Alice McGee

December 2008



Statistics  
Canada

Statistique  
Canada

Canada

## Using behavior coding to evaluate the effectiveness of dependent interviewing

Joanne Pascale and Alice McGee<sup>1</sup>

### Abstract

Dependent interviewing (DI) is used in many longitudinal surveys to “feed forward” data from one wave to the next. Though it is a promising technique which has been demonstrated to enhance data quality in certain respects, relatively little is known about how it is actually administered in the field. This research seeks to address this issue through behavior coding. Various styles of DI were employed in the English Longitudinal Study of Ageing (ELSA) in January, 2006, and recordings were made of pilot field interviews. These recordings were analysed to determine whether the questions (particularly the DI aspects) were administered appropriately and to explore the respondent’s reaction to the fed-forward data. Of particular interest was whether respondents confirmed or challenged the previously-reported information, whether the prior wave data came into play when respondents were providing their current-wave answers, and how any discrepancies were negotiated by the interviewer and respondent. Also of interest was to examine the effectiveness of various styles of DI. For example, in some cases the prior wave data was brought forward and respondents were asked to explicitly confirm it; in other cases the previous data was read and respondents were asked if the situation was still the same. Results indicate varying levels of compliance in terms of initial question-reading, and suggest that some styles of DI may be more effective than others.

Key Words: Dependent interviewing; Longitudinal surveys; Panel surveys; Behavior coding.

### 1. Introduction

In recent years there has been increased interest in and use of “dependent interviewing” (or DI) in longitudinal surveys. DI (also known as “previously reported data” or PRD) is a technique whereby data collected from one wave are carried forward into the next wave in order to tailor question wording and skip patterns. For example, if at Wave 1 a respondent reported working for Employer X, the Wave 2 DI question would read: “Last time you said you worked for Employer X. Are you still working for Employer X?” This is in contrast to an “independent” (that is, non-DI) method whereby at Wave 2 the respondent would simply be asked “from scratch” if he/she was working, and the name of the employer. A related implementation of DI is to route respondents around detailed questions if the circumstances from one wave to another have not changed. For example, a detailed set of questions about Employer X may be asked in Wave 1 (such as the industry, number of employees, *etc.*), and if at Wave 2 the respondent reports they are still working for the same employer, those details need not be collected a second time.

The proliferation of automated surveys has contributed to the increased use of DI, since the technique can be difficult and cumbersome to implement in a paper/pencil questionnaire. Another factor contributing to the interest in DI is its potential to enhance data quality in a number of ways. Generally DI can make for a smoother, smarter, more efficient interview by reminding respondents of their

previous answers and allowing them to simply report whether anything has changed since then. Rigorous research evidence demonstrating this potential is beginning to emerge. For example, there is consistent evidence that DI reduces spurious change, particularly in employment characteristics (Polivka and Rothgeb 1993; Jäckle and Lynn 2004). Another source of measurement error that has consistently plagued panel surveys is “seam bias.”

The “seam” of a panel survey is the point where one wave is joined with the next wave. For example, in a panel survey with annual waves the seam is between December of one year and January of the following year. Seam bias occurs when more transitions (*e.g.*: from employment to unemployment) are observed from December to January than for any of the non-seam month pairs (*e.g.*: February to March, April to May). There is strong evidence that DI significantly reduces (though does not eliminate) this seam bias (Moore, Bates, Pascale and Okon 2006). In terms of respondents’ receptivity to DI, there is qualitative evidence that respondents want and expect it (Pascale and Mayer 2004). In the summer of 2006 a major conference was organized to assess the “state of the art” of research on longitudinal studies and several papers demonstrated specific benefits of DI. An edited monograph book of selected papers is to be published by John Wiley and Sons in 2008 (<http://www.iser.essex.ac.uk/ulsc/mols2006/>).

What the literature seems to lack up to now, however, is evidence of how DI is actually implemented in the field. The current research set out to address this gap. In particular

1. Joanne Pascale, US Census Bureau, 4600 Silver Hill Road, Washington DC 20233. E-mail: joanne.pascale@census.gov; Alice McGee, formerly with the National Centre for Social Research.

we use behavior coding to examine whether interviewers read questions as worded, focusing especially on the dependent words and phrases within the questions, and we examine respondents' reactions to the dependent phrases – that is, whether they affirm or dispute the previously-collected data, and whether providing this information seems to help or hinder the reporting task. Finally we examine whether these behaviors seem to vary at all by “style” of DI – that is, the particular way that the previously-collected information is fed back to the respondent. The vehicle we use for this research is the English Longitudinal Study of Ageing (ELSA), carried out by the National Centre for Social Research (or “NatCen”) in collaboration with University College London and the Institute of Fiscal Studies.

## 2. Methods

### 2.1 ELSA: The survey vehicle

ELSA is a study of people aged 50 and over and their younger partners. The study explores the dynamics of health and disability, family structure, public program participation, economic circumstances, and retirement. The ELSA sample was drawn from households that had previously responded to five years of the Health Survey for England (HSE) between 1998 and 2003. The first ELSA wave was administered in 2002 with 12,100 respondents, and follow-up interviews have been conducted every two years to measure changes in health and social and economic circumstances. Dependent interviewing was embedded in the Wave 2 instrument but due to budget and schedule constraints, little evaluation was done prior to its implementation. Analysis of Wave 2 data, however, raised some concerns about the effect of DI. For example, roughly 20% of respondents who reported high blood pressure at Wave 1 reported that they no longer had the condition at Wave 2. Due in part to this finding, the current research project was undertaken to generally assess the implementation of DI techniques in the field. Behavior coding was chosen as the evaluation method in order to carefully assess interviewer-respondent interactions, and to measure the extent to which the questions were being administered as written.

### 2.2 Field interviewing and recording

The pilot phase of Wave 3 ELSA was conducted over a 4 week period in January, 2006. Altogether 17 NatCen field interviewers from different areas around the United Kingdom conducted 123 individual interviews. The vast majority of the interviews (106) were conducted with individuals who had been interviewed in the prior ELSA wave, while 17 interviews were conducted with members of

a refreshment sample who were new to ELSA but who had been interviewed in the HSE. Most of the analysis in this paper pertains to those individuals interviewed in the prior ELSA wave. However, for two items in the demographics section – LIVE (whether a household member still lives at the residence) and DOB (household member's date of birth) – both the prior-ELSA-wave and the refresher sample were included in the analysis since those items included data fed forward from the HSE interviews. That is, these two items included DI even for those refresher cases not interviewed in ELSA in the prior wave. All interviews were conducted face-to-face using a computer-assisted personal instrument (CAPI). The questionnaire included questions on a number of topics: household and individual demographics, health status, income and assets. Interviews were recorded using Computer Audio Recorded Interviewing (CARI), a software application that allows field interviews to be recorded directly onto computer laptops as digital sound files. A consent question asking respondents for permission to record the interview was embedded into the beginning of the questionnaire, and if respondents did not consent the recorder was not switched on. Furthermore, in some cases the sound files were corrupted and therefore could not be coded. Among the 123 individual interviews, 104 were coded, and among the 106 prior-ELSA-wave interviews, a total of 87 individual recordings were coded. In both cases the majority of the interview losses stemmed from non-consent (versus corrupt sound files).

### 2.3 Dependent interviewing question wording

Dependent interviewing was embedded in the instrument across three different topic areas: demographics, health conditions and vehicle ownership (see Figure 1). In the health condition section there were three broad categories of illnesses: eye, cardiovascular disease (CVD), and chronic conditions. Within each of these broad categories there were multiple specific illnesses asked about. For example under eye conditions there were four illnesses (such as glaucoma and cataracts). Items 4 and 5 were repeated for each illness or condition the respondent had reported in the prior wave.

Five different styles of DI were used across these three topic areas, but as was mentioned earlier, no particular research guided those design decisions. As Figure 1 indicates, each of the six items employed a slightly different style of DI. The first two items in the demographics section (LIVE, DOB) do provide previously-reported data but do not explicitly mention having gathered this data in the previous interview. Rather, the past data is simply presented and the respondent is asked to verify it. The third item (CHILD) explicitly states that the data was collected last time and the respondent is asked if the information is correct. Unlike the demographics questions, the health

questions were separated into two distinct items, which appeared on two different screens. The first (LAST-EYE) was simply a statement, informing the respondent of a particular illness they reported during the previous interview. It was meant to be read as a statement and the respondent was not asked or expected to provide a response to this statement; rather the interviewer was meant to press the “enter” key in order for the second of the two-part series to appear. The second item (STILL-EYE) then asked whether the respondent still had the illness or condition. And finally for vehicle ownership the routine was somewhat similar to the health conditions questions; first a statement was read that informed the respondent of what they reported last time, and then a question was asked to determine if that condition still existed (that is: do you still own the vehicle?). The difference was that for the vehicle item the statement on the past condition and the question (“still have it”) were wrapped into one single item, while in the health section there were two distinct items – the statement and then the “still?” question.

- |  |
|--|
| <p>A. <i>Demographics</i></p> <ol style="list-style-type: none"> <li>1. LIVE: Does NAME still live here?</li> <li>2. DOB: Can I just check, is NAME’s date of birth [fill date of birth (DOB)]?</li> <li>3. CHILD: Our records show that when we last interviewed you, you had a child called NAME, whose date of birth is [DOB]. Are these details correct?</li> </ol> <p>B. <i>Health Conditions</i></p> <ol style="list-style-type: none"> <li>4. LAST-EYE/CVD/CHRON: Our records show that when we last interviewed you in January 2004, you said you had had (or been told by a doctor that you had had) [fill EYE/CVD/CHRONIC CONDITION]. [press enter]</li> <li>5. STILL-EYE/CVD/CHRON: Do you still have [fill EYE/CVD/CHRONIC CONDITION]?</li> </ol> <p>C. <i>Vehicle Ownership</i></p> <ol style="list-style-type: none"> <li>6. VEHICLE: Last time we saw you, you told us that you were the main user of a [MAKE OF VEHICLE], with a [LETTER] registration. Do you still have that vehicle?</li> </ol> |
|--|

**Figure 1** Question wording of items using dependent interviewing

## 2.4 Behavior coding

In order to develop the code frame for behavior coding, we first listened to several recordings to get a general feel for the flow of the interview, the frequency and nature of non-standard interviewer behavior, and respondent’s reactions to the questions. We determined that the “first exchange” – the interviewer’s initial reading of the question and the respondent’s first utterance in response to that – was sufficiently rich for analysis and thus developed a code frame to capture only these behaviors, as well as a final outcome. Within these three behaviors (interviewer’s

initial question-reading, respondent’s initial response, and outcome), we started out with a fairly standard code frame and adapted it based on the content of the recordings and our particular interest in learning about the functioning of the feed-forward phrases embedded within the questions (see Figure 2). For interviewer behavior we used three main code categories: (1) question was read as worded or with only a minor change that did not change the meaning of the question (2) question was read with a “major change” that changed or could change the meaning of the question and (3) the question was omitted. Within the major change code we developed two DI-specific codes. On the recordings it was rather common to hear interviewers changing a statement into a question. For example, in the health section the statement: “Our records show that last time you reported X condition” became a question because interviewers often added “Is that right?” Or, in some cases, interviewers used an intonation and a pause to turn the statement into a question – for example “Our records show that last time you reported X condition?” followed by a pause, waiting for an answer from the respondent. We should note that because coders were working directly from recordings, versus transcripts, they were able to make a judgment regarding the use of intonation to convey either a question or a statement. In other cases a question became a statement. For example in the demographics section the question “Does NAME still live here?” was modified to “And NAME still lives here.” (with no intonation indicating a question mark). Since these were the most frequently-observed problems we created dedicated codes for them.

In total there were 7 coders, drawn from both the survey methods and the operations units. Researchers conducted a half-day training which lasted 4 hours. The training covered the basic concepts of behavior coding, along with the study-specific codes and how to apply them. The majority of training time was devoted to coding hypothetical examples of respondent-interviewer interactions and then discussing and comparing individual judgments and the rationale for those judgments in an attempt to apply codes consistently across coders. However, no formal reliability measures were implemented.

Respondent codes were fairly standard, again with the exception of DI-specific codes. An “adequate” code meant that the respondent’s initial utterance fit into one of the response categories. We adapted this code to capture whether respondents affirmed or disputed the fed-forward data. There were also codes for a request for clarification and a rereading of the question, and a general “inadequate” code, meaning the respondent’s answer did not fit into any of the given response categories. Outcome codes were simply “adequate” and “inadequate.”

<b>A. Interviewer Codes</b>	
S:	Standard; read as worded or with a minor change that did not change the meaning
MC1:	Fed-forward statement was read as a question (e.g.: "Last time you told us you had high blood pressure. Is that correct?")
MC2:	Fed-forward question was read as a statement (e.g.: "And your date of birth was 25 <sup>th</sup> May 1933.")
MC3:	Any other change that did or could change the meaning of the question
O:	Omission
I/O:	Recording was inaudible or the behavior does not fit into one of the above codes
<b>B. Respondent Codes</b>	
AA:	Adequate; acknowledged or did not dispute the fed-forward data
AD:	Adequate; disputed or challenged the fed-forward data
CL:	Request for clarification
IA:	Inadequate answer or elaboration
DK:	Don't know
R:	Refused
I/O:	Recording was inaudible or the behavior does not fit into one of the above codes
<b>C. Outcome Codes</b>	
AA:	Adequate; final response fit one of the given response categories
IA:	Inadequate; final response did not fit any of the response categories
DK:	Don't know
R:	Refused
I/O:	Recording was inaudible or the behavior does not fit into one of the above codes

**Figure 2 Behavior codes**

### 3. Results

Findings will first be presented for each topic area, then themes across topics will be discussed. Regarding the outcome code, adequate answers were obtained in the vast majority of cases (upwards of 90% of the time) and there was little variation across items so those results are not shown.

#### 3.1 Demographic items

In the demographics section, first regarding interviewer behavior, there was fairly wide variation in the extent to which interviewers adhered to standardized technique, ranging from 40-79%, depending on the item (See Table 1A). Note that the base for any given item varies, in this table and others, for two reasons. First, as discussed in Section 2.2, the items LIVE and DOB were administered to both the prior-ELSA-wave sample and the refresher sample (a base of 104 coded interviews), while for all other items only the prior-ELSA-wave sample was used (a base of 87 coded interviews). Second, analysis was conducted at the question level (versus person level) thus some items were administered multiple times for any given individual. For example LIVE was administered for each additional person living in the household, CHILD was administered only if there was a child in the household, and it was repeated for each additional child in the household. Furthermore, the health items were repeated for each illness within a given category. For example, if a respondent had reported six different types of cardiovascular conditions at the prior wave, the health questions would be repeated six times for that individual, once for each illness within the general category of cardiovascular conditions.

**Table 1A**  
**Interviewer behavior for demographic items**

Item	Base (n)	Interviewer Behavior Code (in percent)				
		Read as worded	Question read as statement	Other major change	Omitted	Other
LIVE: Does NAME still live here?	120	40	33	4	18	5
DOB: Can I just check, is NAME's date of birth [DOB]?	107	57	37	1	1	4
CHILD: Our records show that when we last interviewed you, you had a child called NAME, whose date of birth is [DOB]. Are these details correct?	84	79	8	11	0	2

As was predicted from our earlier (unsystematic) listening of the recordings, for the most part when interviewers diverged from the script they turned the question into a statement (e.g.: “Is NAME’s date of birth January 1?” would become “And NAME’s date of birth is January 1.”). This behavior occurred 33-37% of the time for the first two items (LIVE and DOB) and only 8% of the time for the last item (CHILD). This may not be too surprising considering the nature of the items. Answers to the first two items may seem obvious – particularly at Wave 3 – and interviewers may have been somewhat reluctant to ask a question with an obvious answer. Indeed LIVE was omitted altogether 18% of the time, and this could be because the interviewer was talking to the person referenced in the question. The third item, on the other hand, asks about someone else in the household (a child), the information is rather specific (name and date of birth) and the actual question (“are these details correct?”) may not seem to have an obvious answer. That is, it may seem like a more “legitimate” question to ask than asking a person, in what appears to be their home, “Do you still live here?” This could explain why this last item was read as worded so frequently – 79% of the time.

Turning to respondent behavior, on the whole respondents provided a codeable answer straightaway more than 80% of the time (see Table 1B). (Note that in some cases the base for respondent behavior on any given item is smaller than the base for interviewer behavior for that same item. This is due to a combination of interviewers omitting the item (in which case there was no respondent behavior to code) and missing data.) They rarely disputed the fed-forward data (up to only 5% of the time), and most of the disputes stemmed from keying errors in the name or date of the birth previously recorded.

### 3.2 Health items

As noted above, the health questions were asked in two parts. First a statement about the condition reported during the prior wave was read, and then a question was asked to determine whether the condition still existed. Overall levels of “exact reading” of these items were moderate – ranging from 41-76% but generally in the low 60s (see Table 2A). When interviewers diverged from the script they tended to turn the statement into a question (20-38% of the time) by adding something along the lines of “Is that correct?” to the end of the statement. Interviewers would then often omit the actual question “Do you still have it?” altogether – 13-18% of the time. The implications are important here, because it means the respondent is getting a fundamentally different question, specifically “Is it correct that you reported this condition last time?” versus “Do you still have this condition now?”

Another problem was when the actual question “Do you still [have condition X]?” was read, interviewers often read it as a statement rather than a question: “And you still have it.” – 3-16% of the time. This has serious implications for data quality as well, because the respondent is not being given the opportunity to think about whether they really do still have the condition; they are just being told they do.

Regarding respondent behavior (Table 2B), there were fairly high levels of adequate behaviour – over 90% for both cardiovascular and chronic conditions – and 72% for eye conditions (however the base here was only 21 cases). Respondents disputed prior wave data for a variety of reasons. Some said they used to have the condition but no longer do, and this is essentially how the questionnaire was expected to operate. But in other cases the fed-forward data were problematic; respondents either denied that they’d reported the condition at the prior wave, or they disagreed with the characterization of the illness. For example, in one case an illness was recorded as cancer in the prior wave and when asked about it in the next wave the respondent said it wasn’t cancer. He wasn’t sure what the diagnosis was but said it was not cancer. In another case a respondent reported memory impairment at the prior wave but this particular condition was grouped in with other related illnesses in the instrument (“dementia, senility or memory impairment”). When the DI question appeared on the screen the interviewer only read “dementia” and the respondent refuted it. Only when the interviewer went back and read the full question, with all three conditions, did the respondent affirm that he had a memory impairment. And finally, in one case the presence of other household members seemed to be an issue. For example, when the respondent was told he’d reported a certain chronic condition at the prior wave he asked “Did I?” and his wife said “yes.”

### 3.3 Vehicle item

The vehicle item was similar to the health items – first providing a statement about what was recorded in the prior wave and then asking a question about whether the situation is still the same. A key difference, however, was that rather than presenting the statement and question as two distinct items on two different screens, they were rolled into one item. Across all items in the questionnaire the vehicle item had the highest level of interviewers reading the question as worded at 82% (see Table 3A). The problems identified in the health section – interviewers turning the statement into a question, or the question into a statement, or omitting the question – did not turn up very often here, perhaps because the style of DI was different. Specifically, interviewers did not have to have read a statement about the prior report but could move directly into the question: “Do you still have this vehicle?” By not displaying the statement on the prior

wave data as a distinct item, interviewers may have been less tempted to turn that statement into a question by asking, for example, "Is it correct that you reported this vehicle last time?" The result was that the intended question – whether the vehicle was still owned – was being asked, rather than an unintended question ("Did you report owning this vehicle last time?"). However, among the non-standard behaviors there were still several instances of interviewers (8% of the time) turning the question into a statement: "And you still own xx vehicle." This could be a result of interviewers

having seen the vehicle in question on their way to the doorstep.

Respondent behavior here was similar to the health section. Respondents provided a codeable answer straight-away 80% of the time (see Table 3B). They rarely disputed the fed-forward data (6% of the time), and reasons were mixed. One stemmed from keying errors in the fed-forward registration information, and two were based on real change (in one case the respondent had a different car; in the other case the respondent had given up driving).

**Table 1B**  
**Respondent behavior for demographic items**

Item	Base (n)	Respondent behavior code (in percent)				
		Adequate; affirmed fed- forward data	Adequate; disputed fed- forward data	Clarification	Inadequate	Other
LIVE: Does NAME still live here?	91	81	1	1	4	12
DOB: Can I just check, is NAME's date of birth [DOB]?	102	91	0	0	0	9
CHILD: Our records show that when we last interviewed you, you had a child called NAME, whose date of birth is [DOB]. Are these details correct?	84	89	5	1	1	4

**Table 2A**  
**Interviewer behavior for health items**

Item	Base (n)	Interviewer behavior code (in percent)					
		Read as worded	Question read as statement	Statement read as question	Other major change	Omitted	Other
LAST-EYE: Our records show that when we last interviewed you in January 2004, you said you had had (or been told by a doctor that you had had) [condition]	21	62	na	38	0	0	0
STILL-EYE: Do you still have [condition]	19	63	16	na	5	16	0
LAST-CVD: Our records show that when we last interviewed you in January 2004, you said you had had (or been told by a doctor that you had had) [condition]	100	63	na	20	17	0	0
STILL-CVD: Do you still have [condition]	79	76	3	na	8	13	1
LAST-CHRON: Our records show that when we last interviewed you in January 2004, you said you had had (or been told by a doctor that you had had) [condition]	59	41	na	34	17	5	3
STILL-CHRON: Do you still have [condition]	51	61	14	na	2	18	6

**Table 2B**  
**Respondent behavior for health items**

Item	Base (n)	Respondent behavior code (in percent)					
		Adequate; affirmed fed- forward data	Adequate; disputed fed- forward data	Adequate*	Inadequate	Clarification	Other
LAST-EYE: Our records show that when we last interviewed you in January 2004, you said you had had (or been told by a doctor that you had had) [condition]	21	62	10	[72]	10	5	15
STILL-EYE: Do you still have [condition]	16	na	na	94	0	0	6
LAST-CVD: Our records show that when we last interviewed you in January 2004, you said you had had (or been told by a doctor that you had had) [condition]	100	87	5	[93]	4	0	4
STILL-CVD: Do you still have [condition]	67	na	na	69	24	0	8
LAST-CHRON: Our records show that when we last interviewed you in January 2004, you said you had had (or been told by a doctor that you had had) [condition]	53	85	4	[89]	2	0	9
STILL-CHRON: Do you still have [condition]	35	na	na	89	6	0	6

\* For "LAST-XX" items this column shows the sum of "Adequate; affirmed fed-forward data" and "Adequate; disputed fed-forward data"

**Table 3A**  
**Interviewer behavior for vehicle item**

Item	Base (n)	Read as worded	Interviewer behavior code (in percent)			
			Question read as statement	Other major change	Omitted	Other
VEHICLE: Last time we saw you, you told us that you were the main user of a [MAKE OF VEHICLE], with a [LETTER] registration. Do you still have that vehicle?	51	82	8	2	4	4

**Table 3B**  
**Respondent behavior for vehicle item**

Item	Base (n)	Respondent behavior code (in percent)				
		Adequate; affirmed fed- forward data	Adequate; disputed fed- forward data	Inadequate	Clarification	Other
VEHICLE: Last time we saw you, you told us that you were the main user of a [MAKE OF VEHICLE], with a [LETTER] registration. Do you still have that vehicle?	49	74	6	10	4	6



### 3.4 Interviewer-respondent interaction

In addition to item-specific analysis we examined the relationship between interviewer and respondent behavior across all items. We found that whether interviewers read questions as worded, read questions as statements, or read statements as questions, respondents provided an adequate (and affirmative) answer 87.5% of the time. For other behaviors, cell sizes were too small to conduct meaningful analysis.

## 4. Summary and recommendations

The extent to which interviewers adhered to the standardized script varied quite a bit – questions were read as worded 40-82% of the time, depending on the particular item. When interviewers diverged from the script, the way they changed the wording varied by topic area and style of DI which, unfortunately, were confounded because each item had a unique style of DI. In the demographics and vehicle items, for the most part interviewers changed the question into a statement (“Does NAME still live here?” became “And NAME still lives here.”) In the health section interviewers read statements about what was reported last time as questions. Rather than simply reading the statement “Last time you reported X condition” interviewers would add “Is that correct?” (which is an ambiguous question) and often omitted the question “Do you still have condition x?” The result was that often the intended question – to determine whether the condition still exists – was obscured or omitted.

For the most part respondents provided codeable answers on the first exchange 72-94% of the time. It was fairly uncommon for respondents to dispute the fed-forward data (0-10% of the time) but when they did it was for a variety of reasons. Some confirmed the prior wave report but said they no longer have the condition. Some denied the prior wave report, and some disagreed with the details of the fed-forward data. Note that this first scenario is what we expect to happen in the instrument so it is actually a misnomer to say the respondent “disputed” the earlier report. Respondents here are not disputing what they said earlier, but rather they are confirming their earlier report and then reporting change. However, when the code frame was developed we heard very few instances of respondents disputing the prior data at all; the majority of cases were respondents simply agreeing with the fed-forward data. We therefore failed to recognize that it would have been valuable to create separate codes for agreeing to the fed-forward data and reporting real change versus actually disputing the prior report. Even with the full dataset, however, the frequency with which respondents did not simply agree to the prior wave data was too

low for a rich analysis, and a larger dataset would be needed to address this issue.

In terms of recommendations, these findings strongly suggest that questionnaire designers should avoid providing statements of prior wave data without an actual question, because interviewers are too tempted to turn these statements into questions, which obscures the question on whether the prior wave situation still exists. If it is important to confirm or verify information reported in a prior wave, this should be done explicitly by adding a discrete question to the statement, such as: “Last time I recorded that you had condition X. Is that correct?” Subsequent questions could then be asked to determine whether the condition *still* exists. Separating the two concepts in this way would convey to the respondent that there are two distinct issues: one is whether the prior report was recorded accurately, and the other is whether the condition still exists. If researchers do not have a rationale for needing to confirm the accuracy of previously-recorded data, a more efficient approach would be to ask: “Last time I recorded that you had condition X. Do you still have condition X?”

Our findings from the health conditions section suggest that for certain topic areas it is important to feed back prior wave data in the respondent’s own words as much as possible. When respondents’ descriptions of their illnesses was obscured by either the instrument or the interviewer grouping the illness with other conditions, respondents no longer recognized the illness they originally reported.

Finally our findings suggest a more general recommendation that the style of DI should be carefully tailored depending on the particular item. For example, for topics unlikely to change from one wave to the next (such as date of birth), avoid re-asking questions because interviewers often read them as statements or omit them altogether. For these topics it may be more effective to either explicitly verify the accuracy of the earlier report (as suggested above), or to avoid bringing back the information at all. A hybrid-type approach for a study with several waves would be to verify the accuracy of previously-recorded data in wave 2 and then accept the data as correct and avoid re-affirming it in all later waves.

## Acknowledgements

The authors thank Carli Lessof for her ongoing support of survey methods research, and Jeff Moore for reviews of earlier drafts of this paper. We also thank Hayley Cheshire, Michelle Gray, Fiona Andrews, Lizzie Hacker, Colin Miceli, Marian Bolden, Audrey Hale and Laura Street for their assistance with data collection and processing. This research was funded by the National Centre for Social Research, and

funding for the ELSA study is provided by the National Institute of Ageing (in the United States), and a consortium of UK government departments coordinated by the Office for National Statistics.

This report is released to inform interested parties of ongoing research and to encourage discussion. The views expressed on methodological issues are those of the authors and not necessarily those of the US Census Bureau or the National Centre for Social Research.

## References

- Jäckle, A., and Lynn, P. (2004). Dependent Interviewing and Seam Effects in Work History Data. Working Paper #2005-24 of the Institute for Social and Economic Research, Colchester, UK: University of Essex.
- Moore, J.C., Bates, N., Pascale, J. and Okon, A. (2006). Tackling Seam Bias Through Questionnaire Design. Invited paper presented at the conference on Methodology on Longitudinal Surveys, July 12-14, Essex, United Kingdom.
- Pascale, J., and Mayer, T. (2004). Alternative methods for exploring confidentiality issues related to dependent interviewing. *Journal of Official Statistics*, 20, 2, 357-377.
- Polivka, A., and Rothgeb, J. (1993). Redesigning the CPS Questionnaire. *Monthly Labor Review*, September, 10-28.